АНАЛИЗ КРИТЕРИЕВ ОБУЧЕНИЯ НЕЧЕТКОГО КЛАССИФИКАТОРА

ARTICLE · JANUARY 2015

Serhiy Shtovba
Vinnytsia National Technical University

Olga Pankevich
Vinnytsia National Technical University

АВТОМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА. – 2015. – No.3. – С.5-16.

АНАЛИЗ КРИТЕРИЕВ ОБУЧЕНИЯ НЕЧЕТКОГО КЛАССИФИКАТОРА

С.Д. ШТОВБА, *доктор технических наук, профессор*
О.Д. ПАНКЕВИЧ, *кандидат технических наук, доцент*
А.В. НАГОРНАЯ, *аспирант*

Винницкий национальный технический университет
Хмельницкое шоссе 95, Винница 21021, Украина
E-mail: shtovba@gmail.com

В нечетком классификаторе отображение "входы – выход" описывается лингвистическими правилами <Если – то>, антецеденты которых содержат нечеткие термы "низкий", "средний", "высокий" и т.п. Для повышения безошибочности нечеткий классификатор обучают по экспериментальным данным. Рассмотрены задачи с одинаковой и с разной ценой ошибок классификации различных типов. Для задач с неразличимыми типами ошибок в дополнении к двум известным критериям обучения предложен новый. В новом критерии расстояние между желаемым и действительным нечеткими результатами классификации для случаев принятия ошибочного решения взвешивается штрафным коэффициентом. Обобщены критерии обучения для задач классификации с платежной матрицей. Проведенные компьютерные эксперименты на задачах распознавания вин и диагностики заболеваний сердца свидетельствуют, что наилучшие показатели качества настройки нечетких классификаторов достигаются в случае использования нового критерия обучения.

Ключевые слова: обучение, критерии обучения, нечеткие правила, нечеткий вывод, классификация, платежная матрица

1. ВВЕДЕНИЕ

Задача классификации состоит в отнесении объекта к одному из наперед заданных классов. Она осуществляется на основе анализа атрибутов классифицируемого объекта. К классификации сводятся различные задачи принятия решений в инженерии, менеджменте, экономике, политике, медицине, спорте и в других областях.

В последнее время все более популярными становятся нечеткие классификаторы, т.е. классификаторы, в процессе функционирования или обучения которых используются нечеткие множества [1]. Идея применения нечетких множеств для задач классификации впервые изложена в статье [2]. Сегодня наиболее часто применяются классификаторы на основе логического вывода по базе продукционных правил, антецеденты которых содержат нечеткие термы "низкий", "средний", "высокий" и т.п. Каждое правило описывает область факторного пространства, внутри которой объекты принадлежат одному классу. Границы этих областей нечеткие, потому один и тот же объект может одновременно принадлежать нескольким классам, но с разной степенью.

Основные преимущества нечетких классификаторов обусловлены тем, что:

- логический вывод по базе нечетких правил прозрачен – он понятен заказчикам – врачам, экономистам, политикам и специалистам из других прикладных областей с невысокой кибернетической подготовкой;
- модели классификации компактные – для описания сложных разделяющих поверхностей необходимо лишь несколько лингвистических правил;
- формирование базы лингвистических правил обычно не вызывает трудностей у эксперта;
- логический вывод можно реализовать как для числовых, так и для категориальных и нечетких значений входных признаков. При этом в алгоритме логического вывода модифицируется лишь процедура фаззификации [3], а сама модель классификации не изменяется.

Указанные преимущества позволяют нечетким моделям принятия решений успешно конкурировать с классификаторами на основе байесовских правил, метода ближайших соседей, машины опорных векторов, нейронных сетей и других методов индуктивной обработки данных.

Для повышения безошибочности нечеткий классификатор обучают по экспериментальным данным. Существуют 2 подхода к обучению нечеткого классификатора. Первый подход основан на структурной идентификации зависимости “входы–выход” с помощью нечетких правил. Он состоит в формировании базы правил из списка-кандидатов [4], выборе лингвистических квантификаторов, например, “очень”, “более-менее”, для термов антецедентов правил [5] и т.п. Обучение в этом случае сводится к решению дискретной задачи оптимизации. Второй подход основан на параметрической идентификации зависимости “входы–выход” с помощью нечетких правил. При этом во время обучения семантика правил не изменяется, а модифицируются функции принадлежности нечетких термов и весовые коэффициенты правил [1, 6, 7]. Обучение в этом случае сводится к решению задачи оптимизации с непрерывными управляемыми переменными.

В статье рассматривается параметрическая идентификация, во время которой итерационно изменяются параметры классификатора для того, чтобы обеспечить минимальное расстояние между экспериментальными данными и результатами нечеткого вывода. Это расстояние, которое назовем критерием обучения, можно определить по-разному. Цель статьи состоит в выявлении критерия, обучение по которому обеспечит наилучшую безошибочность нечеткого классификатора. Исследуются случаи с одинаковой, и с различной ценой ошибок разных типов. В последнем случае предполагается известной платежная матрица.

2. НЕЧЕТКИЙ КЛАССИФИКАТОР

Обозначим через $\mathbf{X}=(x_1,x_2,\ldots,x_n)$ – вектор информативных признаков (атрибутов) объекта классификации, а через $l_1, l_2,..., l_m$ – классы решений. Тогда нечетким классификатором назовем отображение $\mathbf{X}=(x_1,x_2,\ldots,x_n) \rightarrow y\in\{l_1, l_2,..., l_m\}$, реализуемое с помощью базы нечетких правил. Основываясь на публикациях [1, 4, 6, 7] базу нечетких правил этого отображения запишем так:

$$\text{Если}\left(x_1=\tilde{a}_{1j} \text{ и } x_2=\tilde{a}_{2j} \text{ и ... и } x_n=\tilde{a}_{nj} \text{ с весом } w_j\right), \text{ то } y=d_j, \quad j=\overline{1,k}, \tag{1}$$

где k – количество правил;

$d_j\in\{l_1, l_2, ..., l_m\}$ – значение консеквента *j*-го правила;

$w_j \in [0, 1]$ – весовой коэффициент, задающий достоверность j-го правила, $j = \overline{1, k}$;

$\tilde{a}_{ij}$ – нечеткий терм, оценивающий атрибут x_i в j-ом правиле, $i = \overline{1, n}$, $j = \overline{1, k}$.

Классификация текущего объекта, заданного вектором атрибутов $\mathbf{X}^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$, осуществляется следующим образом. Сначала рассчитывается степень выполнения j-го правила из базы (1):

$$\mu_j(\mathbf{X}^*) = w_j \cdot \left(\mu_j(x_1^*) \wedge \mu_j(x_2^*) \wedge \ldots \wedge \mu_j(x_n^*) \right), \quad j = \overline{1, k}, \tag{2}$$

где $\mu_j(x_i^*)$ – степень принадлежности значения x_i^* нечеткому терму $\tilde{a}_{ij}$;

$\wedge$ – t-норма, которую обычно реализуют операцией минимума или умножением.

Степени принадлежности входного вектора $\mathbf{X}^*$ классам $l_1, l_2, ..., l_m$ рассчитывают так:

$$\mu_{l_s}(y^*) = \underset{\forall j: d_j = l_s}{agg} \left(\mu_j(X^*) \right), \quad s = \overline{1, m}, \tag{3}$$

где agg – агрегирование результатов нечеткого вывода по правилам с одинаковыми консеквентами. Агрегирование реализуем операцией максимума над степенями принадлежности, что соответствует схеме логического вывода с единственным правилом-победителем (*single winner rule*) [8].

Нечетким решением задачи классификации будет нечеткое множество

$$\tilde{y}^* = \left(\frac{\mu_{l_1}(y^*)}{l_1}, \frac{\mu_{l_2}(y^*)}{l_2}, ..., \frac{\mu_{l_m}(y^*)}{l_m} \right). \tag{4}$$

Результатом логического вывода выберем ядро нечеткого множества (4), то есть класс с максимальной степенью принадлежности:

$$y^* = \underset{\{l_1, l_2, ..., l_m\}}{\arg} \; \max_{s = \overline{1, m}} \left(\mu_{l_s}(y^*) \right).$$

Возможна ситуация, когда в ядро нечеткого множества (4) входят несколько элементов. Тогда объект одновременно принадлежит нескольким классам с одинаковыми степенями, значение которых равно $\max_{s = \overline{1, m}} \left(\mu_{l_s}(y^*) \right)$. Для выбора одного из этих конкурентных классов применим схему голосования правил [8]. По этой схеме для каждого из конкурентных классов рассчитаем сумму степеней (2) выполнения соответствующих правил. Решением выбираем класс с максимальной суммой.

3. КРИТЕРИИ ОБУЧЕНИЯ НЕЧЕТКОГО КЛАССИФИКАТОРА БЕЗ УЧЕТА ПЛАТЕЖНОЙ МАТРИЦЫ

Предполагается, что известна обучающая выборка из M пар "входы–выход":

$$\left(\mathbf{X}_r, y_r \right), \quad r = \overline{1, M}, \tag{5}$$

где $y_r \in \{l_1, l_2, ..., l_m\}$.

Введем следующие обозначения:

$\mathbf{P}$ – вектор параметров функций принадлежности нечетких термов из базы правил (1);

$\mathbf{W}$ – вектор весовых коэффициентов правил из базы (1);

$F\left(\mathbf{K}, \mathbf{X}_r\right) \in \{l_1, l_2, ..., l_m\}$ – результат классификации по базе нечетких правил (1) с параметрами $\mathbf{K} = (\mathbf{P}, \mathbf{W})$ для входного вектора $\mathbf{X}_r$ из r-ой строки выборки (5).

Обучение нечеткого классификатора состоит в нахождении вектора $\mathbf{K}$, минимизирующего расстояние между результатами логического вывода и экспериментальными данными из выборки (5). Ниже рассматривается 3 способа расчета этого расстояния, которые назовем критериями обучения.

Критерий 1. Расстояние между желаемым и действительным поведением модели можно определить через частоту ошибок классификации на обучающей выборке:

$$Crit_1 = \frac{1}{M} \sum_{r=\overline{1, M}} \Delta_r(\mathbf{K}), \tag{6}$$

где $\Delta_r(\mathbf{K}) = \begin{cases} 1, & \text{if} \quad y_r \neq F(\mathbf{K}, \mathbf{X}_r) \\ 0, & \text{if} \quad y_r = F(\mathbf{K}, \mathbf{X}_r) \end{cases}$ – ошибка классификации объекта $\mathbf{X}_r$.

Преимущества критерия (6) заключаются в его простоте и ясной содержательной интерпретации. Процент ошибок часто используется как критерий обучения различных систем распознавания образов [9]. В случае критерия (6) целевая функция соответствующей задачи оптимизации принимает дискретные значения, что затрудняет использование быстрых градиентных методов оптимизации, особенно для малых обучающих выборок.

Критерий 2. Качество обучения можно связать с расстоянием между результатом логического вывода в форме нечеткого множества (4) и значением выходной переменной в обучающей выборке. Для этого значение выходной переменной в обучающей выборке (5) преобразуют в такое нечеткое множество [7]:

$$\left.\begin{array}{l} \tilde{y} = \left(\frac{1}{l_1}, \frac{0}{l_2}, ..., \frac{0}{l_m}\right), \quad \text{if} \quad y = l_1 \\ \tilde{y} = \left(\frac{0}{l_1}, \frac{1}{l_2}, ..., \frac{0}{l_m}\right), \quad \text{if} \quad y = l_2 \\ \vdots \\ \tilde{y} = \left(\frac{0}{l_1}, \frac{0}{l_2}, ..., \frac{1}{l_m}\right), \quad \text{if} \quad y = l_m \end{array}\right\}. \tag{7}$$

Критерий обучения на основе расстояния между (4) и (7) записывается следующим образом:

$$Crit_2 = \sqrt{\frac{1}{M} \sum_{r=\overline{1, M}} D_r(\mathbf{K})}, \tag{8}$$

где $D_r(\mathbf{K}) = \sum_{s=\overline{1, m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2$ – расстояние между желаемым и действительным выходными нечеткими множествами при классификации r-го объекта из обучающей выборки (5);

$\mu_{l_s}(y_r)$ – степень принадлежности r-го объекта из обучающей выборки классу l_s согласно (5);

$\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$ – рассчитанная по (3) степень принадлежности классу l_s выхода нечеткой модели с параметрами K в случае входного вектора $\mathbf{X}_r$.

Преимущество критерия $Crit_2$ состоит в учете меры уверенности в принятом решении на основе степеней принадлежности объекта разным классам. В критерии $Crit_1$ эта информация игнорируется, т.е. неважно насколько степень принадлежности у решения больше, чем у других альтернатив – на 0.0001 или на 1. Другими словами, в случае $Crit_1$ считается, что результат классификации объекта является абсолютно достоверным. Кроме того, целевая функция для задачи оптимизации по критерию $Crit_2$ не содержит протяженных плато, поэтому обучать нечеткий классификатор можно с помощью быстрых градиентных методов. Однако, в некоторых случаях [6, 10] оптимальная по (8) нечеткая модель не обеспечивает минимальную безошибочность классификации (6). Это объясняется тем, что близкие к границам классов объекты вносят почти одинаковый вклад D в критерий обучения (8) как при правильной, так и при ошибочной классификации.

Критерий 3. Ниже предлагается новый критерий обучения, который наследует преимущества, рассмотренных выше. Идея состоит в увеличении расстояния D для ошибочно классифицированных объектов:

$$Crit_3 = \sqrt{\frac{1}{M} \sum_{r=\overline{1,M}} \left(\Delta_r(\mathbf{K}) \cdot penalty + 1\right) \cdot D_r(\mathbf{K})}\,, \tag{9}$$

где $penalty > 0$ – штрафной коэффициент.

При $penalty \to 0$ критерии (8) и (9) стают эквивалентными. При $penalty \to \infty$ рельефы целевых функций задач оптимизации на основе критериев (6) и (9) будут схожими. При обучении по критерию $Crit_3$ выбор направления движения к оптимуму в большей мере определяют ошибочно классифицированные объекты. Такое поведение имитирует адаптивный метод оптимизации [11], в котором для повторного обучения, ошибочно распознанные объекты предъявляются чаще. Результаты экспериментов из [11] свидетельствуют, что обучение при таком подходе происходит быстро.

4. КРИТЕРИИ ОБУЧЕНИЯ С УЧЕТОМ ПЛАТЕЖНОЙ МАТРИЦЫ

Платежной называется следующая квадратная матрица:

$$\mathbf{C} = \begin{bmatrix} 0 & c(l_1, l_2) & \dots & c(l_1, l_m) \\ c(l_2, l_1) & 0 & \dots & c(l_2, l_m) \\ \vdots & & & \\ c(l_m, l_1) & c(l_m, l_2) & \dots & 0 \end{bmatrix}, \tag{10}$$

где $c(l_i, l_j)$ – цена ошибки типа $l_i \to l_j$, когда вместо правильного решения l_i в результате классификации ошибочно выбирается решение l_j , $i = \overline{1, m}$, $j = \overline{1, m}$, $i \neq j$. Нули на главной диагонали матрицы (10) указывают на отсутствие платы за правильную классификацию.

При известной платежной матрице (10) критерий 1 преобразуется к виду:

$$Crit_{1C} = \frac{1}{M} \sum_{r=\overline{1,M}} c(u, v)\,, \tag{11}$$

где $u = y_r$ и $v = F(\mathbf{K}, \mathbf{X}_r)$.

Критерий 2 модифицируем таким образом, чтобы слагаемые расстояния $D_r(\mathbf{K})$ были взвешены ценами соответствующих ошибок. В результате формула (8) преобразуется в такую:

$$Crit_{2C} = \sqrt{\frac{1}{M} \sum_{r=\overline{1,M}} D_r(\mathbf{K}, \mathbf{C})}\,, \tag{12}$$

где $D_r(\mathbf{K}, \mathbf{C}) = \sum_{s=\overline{1,m}} \left(1 + c(y_r, l_s)\right) \cdot \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2$ – взвешенное расстояние между желаемым и действительным выходными нечеткими множествами при классификации r-го объекта из обучающей выборки (5).

В критерий 3 вместо $D_r(\mathbf{K})$ введем взвешенное расстояние $D_r(\mathbf{K}, \mathbf{C})$ из (12):

$$Crit_{3C} = \sqrt{\frac{1}{M} \sum_{r=\overline{1,M}} \left(\Delta_r(\mathbf{K}) \cdot penalty + 1\right) \cdot D_r(\mathbf{K}, \mathbf{C})}\,. \tag{13}$$

5. КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ

Целью экспериментов является выявление критерия, обучение по которому обеспечивает наилучшую безошибочность. Рассматриваются 2 тестовые задачи из UCI *Machine Learning Repository* [12]. В первой задаче о распознавании сорта винограда все ошибки классификации равноценны, потому обучение осуществляется без платежной матрицы. Во второй задаче о диагностике заболеваний сердца цена пропуска цели в 5 раз выше, чем ложной тревоги, поэтому при обучении используется соответствующая платежная матрица.

5.1. Задача распознавания сорта винограда

Рассматривается задача определения сорта винограда (y), из которого изготовлено вино. База данных *Wine Dataset* содержит результаты химического анализа по 13-ти показателям 178 образцов итальянских вин, изготовленных в одном и том же регионе. Для каждого образца указан 1 из трех сортов винограда, из которого изготовлено вино.

Обучающую выборку сформируем из строк базы данных с граничными значениями каждого из 13 атрибутов. Дополнительно в обучающую выборку включим все нечетные строки базы данных. Остальные данные занесем в тестовую выборку. В результате получаем обучающую выборку из 100 строк и тестовую – из 78. Спроектируем нечеткий классификатор вина с такими входами: x_1 – alcohol; x_7 – flavanoids и x_{13} – proline. Воспользуемся базой правил (табл.1) нечеткого классификатора вина из статьи [13]. Нечеткие термы зададим гауссовою функцией принадлежности с двумя параметрами:

$$\mu(x) = \exp\left(-\frac{(x-b)^2}{2c^2}\right), \tag{14}$$

где b – координата максимума и $c > 0$ – коэффициент концентрации.

Параметры функций принадлежностей исходного нечеткого классификатора приведены в табл.2.

Для каждого критерия проведем 500 экспериментов по обучению нечеткой модели на основе квазиньютоновского алгоритма. После обучения каждый классификатор проверим на тестовой выборке по критерию $Crit_1$. Эксперименты проведем для двух нечетких моделей, в одной из ко-

Таблица 1.

База нечетких правил классификатора вина

№	x_1	x_7	x_{13}	y
1	–	–	Низкий	Сорт 1
2	Низкий	–	–	Сорт 2
3	–	Низкий	–	Сорт 3

Таблица 2.

Параметры функций принадлежности термов нечетких классификаторов вина

Атрибут	Терм	Исходный		Best_min		Best_prod	
		b	c	b	c	b	c
x_1	Низкий	11	1.65	11	1.05	11	1.11
x_7	Низкий	0.34	2	0.34	0.91	0.34	0.764
x_{13}	Низкий	2.78	6	2.78	10	2.78	10

торых t-норма реализована операцией минимума, а в другой – произведением. Во время обучения настроим весовые коэффициенты каждого из 3-х правил базы знаний и коэффициенты концентрации (c) функций принадлежности каждого нечеткого терма. В базе знаний все нечеткие термы являются крайними, поэтому согласно [14] координаты максимумов функций принадлежности (коэффициенты b) настраивать не будем, а приравняем к минимальному значению соответствующего атрибута. Таким образом, общее количество настраиваемых параметров составляет 3+3=6. Начальные точки для обучения выбирались случайно – для весовых коэффициентов правил из диапазона [0, 1], а для параметров функций принадлежности в пределах $\pm 20\%$ от значений табл.2.

Вначале определим приемлемые значения штрафного коэффициента в критерии $Crit_3$. Для этого для каждой нечеткой модели проведем по 500 экспериментов, в которых штрафной коэффициент выбирался случайно из интервала (0, 10]. Затем разобьем этот интервал на 5 отрезков равной длины и для каждого из них подсчитаем частоту успешных экспериментов (α). Успешным будем считать эксперимент, результат которого попадает в 25% лучших по уровню безошибочности на тестовой выборке. В ходе экспериментов установлено (рис.1), что наибольшее количество успешных запусков наблюдается когда $penalty \in (0, 2]$. Такие значения штрафного коэффициента будем использовать в дальнейших исследованиях.

После серии из 3000 экспериментов мы получили множество из 10 классификаторов, каждый из которых обеспечивает на тестовой выборке минимальную частоту ошибок (6) на уровне 0.039. Из этого множества выберем 2 классификатора Best_min и Best_prod с минимальными значениями этого критерия и на обучающей выборке, а именно $Crit_1(\text{Best\_min}) = 0.1$ и $Crit_1(\text{Best\_prod}) = 0.1$. Здесь и далее через Best_min обозначен лучший классификатор, у которого t-норма реализована операцией минимума, а через Best_prod – лучший классификатор, у которого t-норма реализована произведением. Весовые коэффициенты правил в Best_min равны $\omega_1 = 0.26$, $\omega_2 = 1$, $\omega_3 = 0.49$, а у Best_prod – $\omega_1 = 0.29$, $\omega_2 = 1$, $\omega_3 = 0.72$. Параметры функций принадлежности этих классификаторов сведены в табл.2.

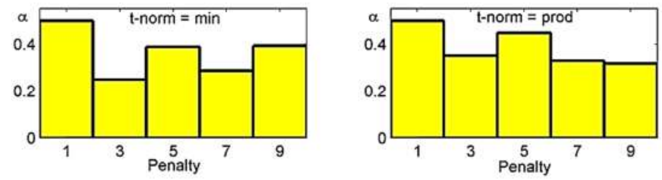

Рис.1. Зависимость успешности обучения нечеткого классификатора вина от штрафного коэффициента.

Таблица 3.

Статистика обучения нечетких классификаторов вина

t-норма	Критерий обучения	Значение критерия $Crit_1$ на тестовой выборке			
		минимальное	среднее	максимальное	СКО
минимум	$Crit_1$	0.103	0.582	0.962	0.143
	$Crit_2$	0.051	0.229	0.564	0.058
	$Crit_3$	0.039	0.198	0.513	0.114
произведение	$Crit_1$	0.115	0.599	0.974	0.142
	$Crit_2$	0.09	0.223	0.449	0.049
	$Crit_3$	0.039	0.2	0.68	0.124

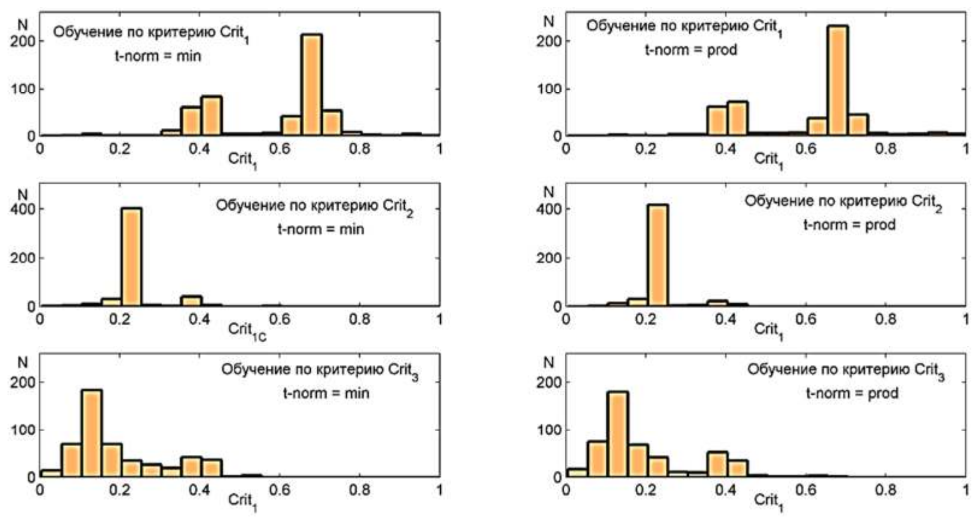

Рис.2. Распределение результатов обучения нечетких классификаторов вина (N – количество случаев).

Результаты экспериментов (табл.3 и рис.2) свидетельствуют, что в среднем лучшее качество обучения наблюдается при использовании критерия $Crit_3$. Самый широкий разброс результатов обучения наблюдается при оптимизации по критерию $Crit_1$, а самый узкий – в случае критерия $Crit_2$.

Создадим модель определения сорта винограда с помощью альтернативного метода, которым выберем дерево решений. С использованием правил расщепления на основе индекса Джини, энтропии и половинного деления (*twoing rule*) [15] синтезировано 3 дерева решений. После подрезания наилучшим оказалось дерево на основе энтропийного расщепления. Дерево содержит 7 правил с длиной антецедентов от 2 до 4 логических условий. Частота ошибок этого дерева на тестовой выборке составляет 0.103. Следовательно, для задачи определения сорта винограда нечеткий классификатор по критериям безошибочности и компактности оказался лучше дерева решений.

5.2. Задача диагностики заболеваний сердца

Рассматривается задача диагностики болезни сердца по данным из *Statlog Heart Data Set*. В каждой из 270 строк этой базы данных описано 13 признаков состояния пациента, по которым принимается решение (y) о наличии или отсутствии болезни сердца. Известна следующая платежная матрица: $\mathbf{C}=\begin{bmatrix}0 & 1\\ 5 & 0\end{bmatrix}$.

Обучающую выборку сформируем из строк базы данных, содержащих граничные значения каждого из 13 атрибутов. Дополнительно в обучающую выборку включим все нечетные строки базы данных. Остальные данные занесем в тестовую выборку. В результате получаем обучающую выборку из 145 строк и тестовую – из 125.

Спроектируем нечеткий классификатор с тремя входами:

x_1 – age; x_{10} – old peak; x_{12} – number of major vessels colored by fluoroscopy.

По распределению данных сформирована нечеткая база знаний (табл.4). Нечеткие термы зададим гауссовой функцией принадлежности (14) с параметрами из табл.5.

Таблица 4.

База нечетких правил для диагностики заболеваний сердца

№	x_1	x_{10}	x_{12}	y
1	–	Низкий	Мало	Здоровый
2	Молодой	Высокий	Мало	Здоровый
3	–	–	Много	Больной
4	Старый	Высокий	Мало	Больной

Таблица 5.

Параметры функций принадлежности термов нечетких классификаторов для диагностики заболеваний сердца

Атрибут	Терм	Начальный		Best_min		Best_prod	
		b	c	b	c	b	c
x_1	Низкий	29	20.4	29	10.4	29	8.75
	Высокий	77	20.4	77	8.69	77	8.38
x_{10}	Низкий	0	2.63	0	0.83	0	0.93
	Высокий	6.2	2.63	6.2	3	6.2	1.11
x_{12}	Низкий	0	1.27	0	0.57	0	0.45
	Высокий	3	1.27	3	2	3	2

Во время обучения настроим весовые коэффициенты каждого из 4-х правил базы знаний и коэффициенты концентрации (c) функции принадлежности каждого нечеткого терма. В базе знаний все нечеткие термы являются крайними, поэтому координаты максимумов функций принадлежности приравняем к границам диапазона изменения атрибутов. Таким образом, общее количество настраиваемых параметров составляет 4+6=10. Начальные точки для обучения выбирались случайно – для весовых коэффициентов правил диапазона [0, 1], а для параметров функций принадлежности в пределах $\pm 20\%$ от значений табл.5.

Как и для предыдущей задачи, вначале определим приемлемые значения штрафного коэффициента в критерии $Crit_{3C}$. В результате 500 экспериментов установлено (рис.3), что наибольшее количество успешных запусков наблюдается когда $penalty \in (0, 2]$. Такие значения штрафного коэффициента будем использовать в дальнейших экспериментах.

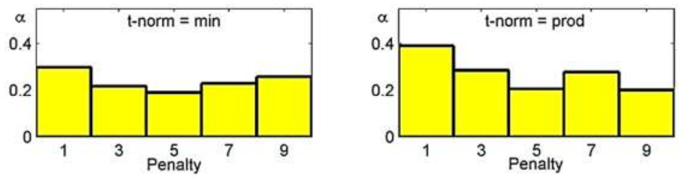

Рис.3. Зависимость успешности обучения нечеткого классификатора для диагностики заболеваний сердца от штрафного коэффициента.

После серии из 3000 экспериментов мы получили множество из 8 классификаторов, каждый из которых обеспечивает на тестовой выборке минимальное значение критерия $Crit_{1C}$ на уровне 0.424. Из этого множества выбираем 2 классификатора Best_min и Best_prod с минимальными значениям этого критерия и на обучающей выборке, а именно $Crit_{1C}(\text{Best\_min}) = 0.683$ и $Crit_{1C}(\text{Best\_prod}) = 0.683$. Весовые коэффициенты правил в Best_min равны $\omega_1 = 0.75$, $\omega_2 = 0.24$, $\omega_3 = 0.96$ и $\omega_4 = 1$, а у Best_prod – $\omega_1 = 0.67$, $\omega_2 = 0.63$, $\omega_3 = 1$ и $\omega_4 = 0.27$. Параметры функций принадлежности этих классификаторов сведены в табл.5.

Результаты экспериментов (табл.6 и рис.4) свидетельствуют, что в среднем лучшее качество обучения наблюдается при использовании критерия $Crit_{3C}$. Как и для предыдущей задачи самый широкий разброс результатов обучения наблюдается при оптимизации по критерию $Crit_1$, а самый узкий – в случае критерия $Crit_2$.

Таблица 6.

Статистика обучения нечетких классификаторов для диагностики заболеваний сердца

t-норма	Критерий обучения	Значение критерия $Crit_{1C}$ на тестовой выборке			
		минимальное	среднее	максимальное	СКО
минимум	$Crit_{1C}$	0.424	1.185	1.408	0.309
	$Crit_{2C}$	0.96	1.081	1.112	0.015
	$Crit_{3C}$	0.424	0.592	1.032	0.094
произведение	$Crit_{1C}$	0.424	1.051	1.408	0.399
	$Crit_{2C}$	0.832	0.923	0.952	0.019
	$Crit_{3C}$	0.424	0.59	1.288	0.103

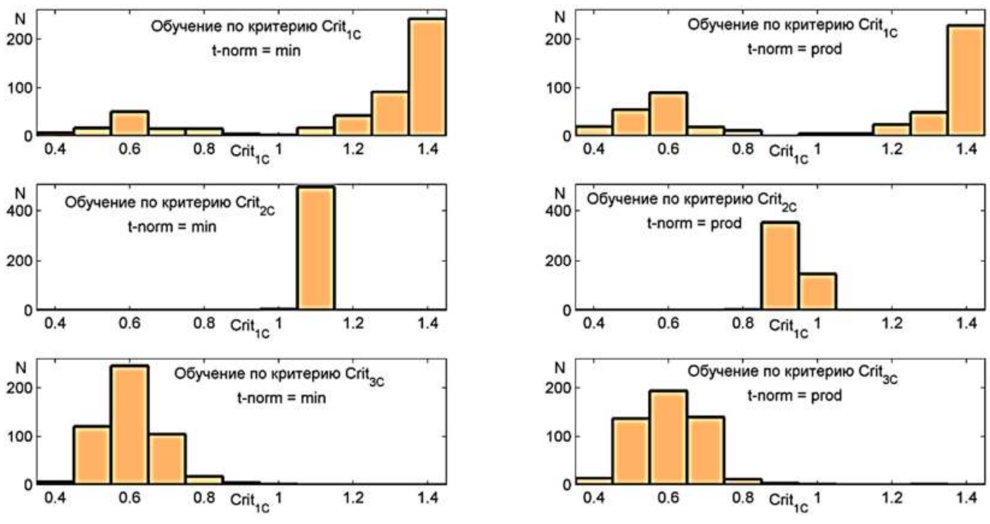

Рис.4. Распределения результатов обучения нечетких классификаторов
для диагностики заболеваний сердца (*N* – количество случаев).

Сравним нечеткий классификатор с альтернативными моделями на основе дерева решений. Из трех моделей, синтезированных с использованием правил расщепления на основе индекса Джини, энтропии и половинного деления, после подрезания наилучшим оказалось дерево на основе энтропийного расщепления. Дерево содержит 12 правил, в каждом из которых от 3 до 4 логических условий. Риск (11) этого дерева на тестовой выборке составляет 0.424, т.е. такой же, как и для нечеткого классификатора. Однако нечеткая модель получилась значительно компактнее дерева решений.

6. ЗАКЛЮЧЕНИЕ

Рассмотрены классификаторы на основе базы нечетких правил для задач с одинаковой и с разной ценами ошибок классификации различных типов. Для задач с неразличимыми типами ошибок, в дополнении к двум известным критериям обучения нечеткого классификатора на основе частоты ошибок и расстояния между нечеткими множествами, предложен новый. В новом критерии расстояние между желаемым и действительным нечеткими результатами классификации в случае принятия ошибочного решения взвешивается штрафным коэффициентом. Обобщены критерии обучения для задач, в которых цены ошибок классификации заданы платежной матрицей. Проведенные компьютерные эксперименты на задачах распознавания вина и диагностики заболеваний сердца свидетельствуют, что наилучшие показатели качества настройки нечетких классификаторов достигаются в случае использования предложенного критерия обучения.

СПИСОК ЛИТЕРАТУРЫ

[1] *Kuncheva L.I.* Fuzzy classifier design / Studies in Fuzziness and Soft Computing. Vol. 49. – Berlin – Heidelberg: Springer-Verlag, 2000.

[2] *Bellman R., Kalaba R., Zadeh L.* Abstraction and pattern classification // Journal of Mathematical Analysis and Applications. – 1966. Vol. 13, №1. P. 1–7.

[3] *Rotshtein A., Shtovba S.* Identification of a nonlinear dependence by a fuzzy knowledgebase in the case of a fuzzy training set // Cybernetics and Systems Analysis. – 2006. – Vol. 42, №2. – P.176–182.

[4] *Ishibuchi H., Nakashima T., Nii M.* Classification and modeling with linguistic information granules: advanced approaches advanced approaches to linguistic data mining. – Berlin – Heidelberg: Springer-Verlag, 2005.

[5] *Cordon O., del Jesus M.J., Herrera F.* Genetic learning of fuzzy rule-based classification systems cooperating with fuzzy reasoning methods // International Journal of Intelligent Systems. – 1998. Vol. 13, №10–11. P. 1025–1053.

[6] *Штовба С.Д.* Проектирование нечетких систем средствами MATLAB. М.: Горячая линия – Телеком, 2007.

[7] *Rotshtein A.* Design and tuning of fuzzy rule-based system for medical diagnosis. In “Fuzzy and Neuro-Fuzzy Systems in Medicine” (Eds.: Teodorescu N.H., Kandel A., and Jain L.C.). Boca–Raton: CRC–Press, 1998.P. 243–289.

[8] *Ishibuchi H., Nakashima T., Morisawa T.* Voting in fuzzy rule-based systems for pattern classification problems // Fuzzy Sets and Systems. – 1999. Vol. 103, №2. P. 223–238.

[9] *Дуда Р., Харт П.* Распознавание образов и анализ сцен. М.: Мир, 1976.

[10] *Shtovba S., Pankevich O., Dounias G.* Tuning the fuzzy classification models with various learning criteria: the case of credit data classification // Proc. of Inter. Conference on Fuzzy Sets and Soft Computing in Economics and Finance. St. Petersburg (Russia), 17–20 June 2004. Vol.1. – St. Petersburg: Russian Fuzzy Systems Association, 2004. P. 103–110.

[11] *Растригин Л.А.* Адаптация сложных систем. Методы и приложения. Рига: Зинатне, 1981.

[12] *Bache K., Lichman M.* UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science. – 2013.

[13] *Ishibuchi H., Yamamoto T.* Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining // Fuzzy Sets and Systems. – 2004. Vol. 141, №1. P. 59–88.

[14] *Shtovba S.* Ensuring accuracy and transparency of Mamdani fuzzy model in learning by experimental data // Journal of Automation and Information Sciences. – 2007. Vol. 39, №8.P. 39–52.

[15] *Breiman L.* Technical note: some properties of splitting criteria // Machine Learning. – 1996. – Vol.24. – №1. – P. 41–47.

Рукопись получена 04.09.2014,
финальная версия – 02.02.2015.

Словарь специфических терминов для переводчика

нечеткое множество - fuzzy set; нечеткий вывод – fuzzy inference;
машина опорных векторов – support vector machine; квантификатор – hedge;
терм - term; функция принадлежности – membership function;
платежная матрица – cost matrix; ядро – core; обучающая выборка – training set;
тестовая выборка – test set; индекс Джини – Gini index;
дерево решений – decision tree.